

Comparison of CNNs and SVM for voice control wheelchair

Mohammad Shahrul Izham Sharifuddin, Sharifalillah Nordin, Azliza Mohd Ali

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

Article Info

Article history:

Received Feb 10, 2020

Revised Apr 20, 2020

Accepted May 6, 2020

Keywords:

Convolutional neural networks

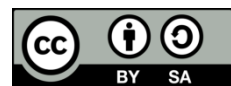
Support vector machine

Voice recognition

ABSTRACT

In this paper, we develop an intelligent wheelchair using CNNs and SVM voice recognition methods. The data is collected from Google and some of them are self-recorded. There are four types of data to be recognized which are go, left, right, and stop. Voice data are extracted using MFCC feature extraction technique. CNNs and SVM are then used to classify and recognize the voice data. The motor driver is embedded in Raspberry PI 3B+ to control the movement of the wheelchair prototype. CNNs produced higher accuracy i.e. 95.30% compared to SVM which is only 72.39%. On the other hand, SVM only took 8.21 seconds while CNNs took 250.03 seconds to execute. Therefore, CNNs produce better result because noise are filtered in the feature extraction layer before classified in the classification layer. However, CNNs took longer time due to the complexity of the networks and the less complexity implementation in SVM give shorter processing time.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Sharifalillah Nordin,
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia.
Email: sharifa@fskm.uitm.edu.my

1. INTRODUCTION

Impaired people is increasing every year due to ageing, accidents and also diseases such as paralysis, spinal cord injuries (SCI), amputation and quadriplegia [1]. This group of people require a special wheelchair to move such as a motorised wheelchair. Motorised wheelchair consists of electric motor and joystick at the armrest. However, disabled people with paralysed hand or having a problem with the hand cannot operate joystick on the motorised wheelchair to move. Therefore, a lot of researches studied to help the impaired people move the wheelchair-using voice recognition [1-3]. This wheelchair can be called smart or intelligent wheelchair because it can be moved using only the voice. The first intelligent wheelchair was developed by the Siamo University of Alcala, Spain in 1999 [4].

The improvement of voice and speech identification technology has been started by Texas Instruments in 1960 [5]. Recently, voice recognition or speech recognition has been applied in assisting people doing work through digital devices such as mobile phone, tablets, and personal computer. There are also voice recognition software and webpages i.e. google applications, translation software, and personal assistants such as Alexa, Cortana and Siri. This personal assistant can be called modern chatbot which can assist people in any topics they would like to ask.

The advancement of technology in artificial intelligence (AI) have created a smart wheelchair or intelligent wheelchair exploiting the voice recognition features. A lot of researches have been conducted to improve the functionality of the wheelchair. Nasrin [1] proposed an application on smart wheelchair-using voice and add GPS function to track the user navigation and location. This application required WIFI connection and the application is installed in a mobile environment. However, Avutu et. al. [3] proposed a

low-cost map for voice recognition wheelchair which can be applied in a local location. This application can be used without any network connection. The cost is much cheaper compared to the application using a network connection. Barriuso [6] proposed a smartphone application for a wheelchair with an agent-based intelligent interface. Another application was developed using Arduino Mega 2560 by [7]. The special features in this application are the emergency messages can be sent to the important people. With the ultrasonic sensor, any obstacle can be detected and avoided without internet connection required. There is also an application of intelligent wheelchair for people with paraplegia problem proposed by [8]. The application applied voice recognition and touch screen control. Meanwhile, Chauhan [9] did a comparative study on voice recognition wheelchair and proposed a new design deploying infrared sensor and Raspberry Pi through USB microphone.

The development of smart or intelligent wheelchair mostly based on AI technologies. The process of voice recognition application will begin extracting features and classification. There are a lot of features extraction technique for voice recognition such as Linear Prediction Coefficients (LPC), Discrete Wavelet Transform (DWT), Line Spectral Frequencies (LSF), and Mel-Frequency Cepstrum Coefficients (MFCC) [10].

Classification techniques can be applied in classifying and recognising the voice, especially in wheelchair application. The most popular classification techniques and promising with high accuracy are convolutional neural networks (CNNs). CNNs is based on feedforward neural network and have better generalisation than networks with full connectivity between adjacent layers [11]. Moreover, CNNs has been successfully applied in many applications especially in image identification [12-15], surveillance [16], and human recognition [17-19]. Lei and She [20] authenticate voice using CNNs method in a noisy environment. Their research produces better accuracy and reduces an equal error rate. Guan [21] optimise performance in speech recognition using CNNs and the recognition performance has increased with an error rate of 13.88%. Sharifuddin et.al compared CNNs and BPNNs on voice control wheelchair applications and produces high accuracy on CNNs technique [5].

Support Vector Machine (SVM) can be categorised as one of the best classifications techniques. SVM also has been applied in various applications for instance in biometrics [17, 22-23], sentiment analysis [24-25] and security such as intrusion detection [26-27]. Selvakumari and Radha [28] applied SVM in classifying speech pathology and achieved 98% accuracy compared to the Naïve Bayes algorithm. Furthermore, Astuti and Riyandwita [29] applying SVM in recognising voice for starting a car engine. SVM classified the word 'on' and 'off' car engine with 92.15% accuracy. Meanwhile, Harvianto et. al [30] analyse the voice of Indonesia Language using MFCC and SVM achieved high accuracy which is 91.83%.

In this paper, we propose an intelligent wheelchair-using voice recognition. There are four types of voice command to be recognized which are go, left, right and stop. The data is collected from Google and some of them are self-recorded. Features from the data are extracted using MFCC techniques. Followed by the recognition using CNNs. We also classify the voice using the SVM method to compare the efficiency and accuracy between the two techniques. This paper is organised as follows. Section 2 presents research methods on voice recognition, CNNs and SVM. Experimentation details and results with the comparisons between CNNs and SVM are presented in section 3. Section 4 discusses the conclusion of the paper.

2. RESEARCH METHOD

This paper focuses on five types of voice data. Speech signal is based on single words; speakers are independent; language is English; vocabulary type is small, and background noise is true. Table 1 presents the five types of voice data used in this paper.

Table 1. Types of collected voice data

Type	Description
Speech signal	Single words
Speaker	Individual
Language	English
Vocabulary	Small
Background noise	Yes

Development board contains microprocessor and microcontroller in a small, compact circuit board. The development board receive the input, process the command, and produce the output. Several development boards are in the market with their strengths and limitations, for example Arduino, Banana PI, and Raspberry

PI. Raspberry PI 3B+ will be used in this project due to the following reasons; fast processing time, low cost and easy to program.

In voice recognition, CNNs is seen as a technique capable of providing high accuracy in solving especially in speech and image recognition problems [20]. CNNs is good in highlighting important points, as well as filtering noise in the dataset. CNNs consist of feature extraction layer and classification layer.

2.1. Feature extraction layer

Feature extraction layer contains a convolutional layer and pooling layer. In this section, data will be extracted before been feed into the classification layer. We implement 2D Convolutional Layer and three 2D Max Pooling layers in this paper.

2.2. Classification layer

Classification layer aims to classify the data received from the feature extraction layer. In this paper, the classification layer consists of an input layer, a hidden layer, and an output layer.

On the other hand, this paper also highlights the implementation of SVM to classify the extracted voice data. SVM is chosen because of the ability to perform multi-class classification with high dimensional spaces. Three kernel functions are used in the experiment, i.e. polynomial, RBF, and sigmoid. We then conduct another experiment to calculate the effect of C, gamma, and degree using the kernel result.

Figure 1 presents the flow of the proposed system [5]. We described the flow based on three modules: the input (the user i.e. the disabled people), the process, and the output. Firstly, the user pushes the button to record the voice command using the portable microphone. MFCC and CNNs are used in the Raspberry PI 3B+ to process the recorded voice. The output signal is then sent to the motor driver in the Raspberry PI 3B+ after the system recognized the voice. The motor driver controls the movement of the motor to direct the robot car.

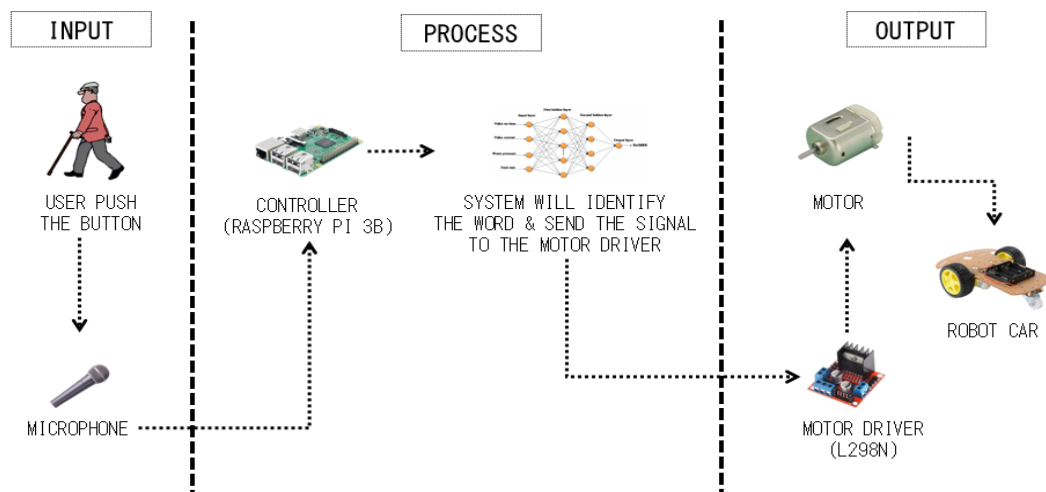


Figure 1. Flow of the system

3. EXPERIMENTATIONS AND RESULTS

3.1. Data Sample

There are four types of voice command used in the experiment. These commands are right, left, stop, and go. The data was gathered from Google. There are 2,372 data for go command, 2,380 for stop command, 2,353 for left command, and 2,367 for right command. There are also urban noise and white noise data collected for the experiment. 2,373 urban noise data are collected from Google while 2,300 white noise data are self-recorded [5] because the data is not available online. Each voice data is prepared in one-second length and is saved in WAV format. The total number of downloaded and recorded voice command used in the experiment is presented in Table 2.

Table 2. Total number of collected voice data

Types	Data
Go	2,372
Stop	2,380
Left	2,353
Right	2,367
Urban Noise	2,373
White Noise	2,300
Total	14,145

3.2. Mel frequency cepstral coefficients (MFCC)

MFCC is the technique used to extract features from audio signal. MFCC used the frequency band logarithmically, which allow better speech processing [31]. Signal features are first went through feature extraction and later feature matching. The processing of MFCC is applied using Librosa library. There are six steps involve in MFCC features extraction i.e. pre-emphasis filter, framing the speech sample, windowing, fast fourier transform (FFT), mel filter bank, and log energy and discrete cosine transform (DCT).

3.3. Data preparation

This research using 2-dimensional (2D) CNNs. The data is directly taken from raw MFCC because the dataset is in 2D format and no requirement to compress it. The size of the data has to be in the same format which is (44, 20). This format is being set based on one-second length of data. If the size of the data exceeded, then the excess data will be cut off. However, if the size is less than the format size (44,20) then it will be pad with zero. Finally, the data is ready to be feed in the 2D CNNs. Figure 2 illustrates the data preparation process of CNNs [5].

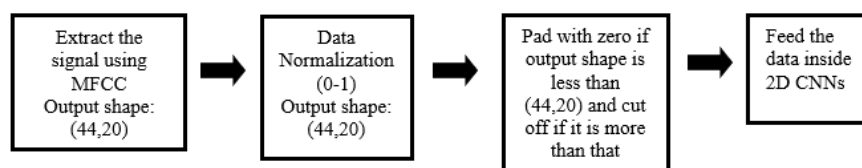


Figure 2. CNNs data preparation

3.4. SVM parameter tuning

Support Vector Machine (SVM) is another technique used to classify the voice data. In order to fine-tune the SVM, there are fourteen experiments have been conducted. The parameter used is in this experiment is a type of kernel, C which is regularization parameter, gamma and degree. We conducted three experiments using three different kernel functions i.e. polynomial, RBF and sigmoid. The result shown that the polynomial kernel achieved the higher accuracy. Next, we conducted another experiments to measure the effect of C continue with the effect of gamma and the effect of degree. We achieved the best accuracy result that is 72.39% with kernel type polynomial, the effect of C is 10, the effect of gamma is scale, and the effect of degree is 1 as shown in Table 3.

Table 3. SVM best model

Kernel	C	Gamma	Degree	Accuracy
Poly	10	Scale	1	72.39%

3.5. CNNs parameter tuning

Twelve experiments were conducted to tune CNNs parameter. From the experiments, we described the best-selected model in Table 4 [6]. Table 4 describes the layers, kernel size, stride, filters, padding, nodes, bias, and activation for the CNNs model. We achieved 95.30% accuracy with 4.1672 minutes from the implemented model.

Table 4. The best model of CNNs

Layers	Kernel Size	Stride	Filters	Padding	Nodes	Bias	Activation
Input				20,44,1			
2D Convolutional	2	1	16	valid	-	False	Relu
2D Max Pooling	2	2	-	valid	-	-	-
2D Convolutional	2	1	32	valid	-	False	Relu
2D Max Pooling	2	2	-	valid	-	-	-
2D Convolutional	2	1	64	valid	-	False	Relu
2D Max Pooling	2	2	-	valid	-	-	-
Flatten		-			-		-
Output		-			6	False	Softplus
Optimizer				Adam			
Loss Function				mean_squared_error			
Epochs				50			
Batch Size				10			

3.5. Comparison between SVM and CNNs

Based on the conducted experiment, CNNs achieved higher accuracy that is 95.30% with compared to SVM that is 72.39%. In term of time, SVM took shorter processing time with only 8.21 second rather than CNNs with 250.03 seconds to run. The capability of filtering the noise in the voice data in the CNNs feature extraction layer produces higher accuracy compared to SVM. On the other hand, the less complexity implementation in SVM give shorter processing time. Table 5 displays the accuracy and time result for CNNs and SVM.

Table 5. Comparison of results

Model	Accuracy	Time (s)
SVM	72.39%	8.21
CNNs	95.30%	250.03

3.6. Hardware tuning

To demonstrate the implementation of the voice recognition process, we integrate logic gates in the Motor Driver LN298N. The description of the logic gates is shown in Table 6. The logic gates control the movement of the motor in four directions, i.e. go, stop, left, and right [6].

Table 6. Implementation of motor driver

Motor Driver Pin LN298N				Motor Direction
In1	In2	In3	In4	
0	0	0	0	Stop
0	1	0	1	Go
1	0	0	1	Left
0	1	1	0	Right

4. CONCLUSION

In this paper, we propose an intelligent wheelchair-using voice recognition. There are four types of voice command to be recognized which are go, left, right and stop. The data is collected from Google and some of them are self-recorded. Features from the data are extracted using MFCC techniques. Followed by the recognition using CNNs and SVM. CNNs produced higher accuracy i.e. 95.30% compared to SVM which is only 72.39%. On the other hand, SVM only took 8.21 seconds while CNNs took 250.03 seconds to execute. Therefore, CNNs produce better result because noise are filtered in the feature extraction layer before classified in the classification layer. However, CNNs took longer time due to the complexity of the networks and the less complexity implementation in SVM give shorter processing time.

ACKNOWLEDGEMENTS

The authors would like to thank the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia for the support throughout this research.

REFERENCES

- [1] N. Aktar, I. Jahan, and B. Lala, "Voice Recognition based Intelligent Wheelchair and GPS Tracking System," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019, pp. 7-9.
- [2] M. A. A. Azmin and S. Nordin, "Voice Control For Wheelchair Movement: A Simulation," *Universiti Teknologi MARA*, 2016.
- [3] S. R. Avutu, "Voice control module for Low cost Local-Map navigation based Intelligent wheelchair," *2017 IEEE 7th Int. Adv. Comput. Conf.*, pp. 609-613, 2017.
- [4] R. C. Simpson, "Smart wheelchairs: A Literature Review," *J. Rehabil. Res. Dev.*, vol. 42, no. 4, p. 423, 2005.
- [5] M. S. Sharifuddin, S. Nordin, and A. Mohd Ali, "Voice Control Intelligent Wheelchair Movement Using CNNs," in *2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, pp. 40-43, 2019.
- [6] D. L. Barriuso, J. Pérez-Marcos, D. M. Jiménez-Bravo, G. V. González, and J. F. De Paz, "Agent-Based Intelligent Interface for Wheelchair Movement Control," *Sensors*, vol. 18, pp. 1-31, 2018.
- [7] Z. Raiyan, "Design of an Arduino Based Voice-Controlled Automated Wheelchair," pp. 21-23, 2017.
- [8] C. Aruna, A. D. Parameswari, M. Malini, and G. Gopu, "Voice Recognition and Touch Screen Control Based Wheel Chair for Paraplegic Persons," *2014 Int. Conf. Green Comput. Commun. Electr. Eng.*, pp. 1-5.
- [9] S. Chauhan, a. S. Arora, and A. Kaul, "A survey of emerging biometric modalities," *Procedia Comput. Sci.*, vol. 2, pp. 213-218, 2010.
- [10] S. Alim and A. K. Rashid, "Some Commonly Used Speech Feature Extraction Algorithms," *IntechOpen*, pp. 2-19, 2018.
- [11] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nat. Int. Wkly. J. Sci.*, vol. 521, pp. 436-444, 2015.
- [12] S. B. Jadhav, V. R. Udupi, S. B. Patil, and D. Cnn, "Convolutional neural networks for leaf image-based plant disease classification," vol. 8, no. 4, pp. 328-341, 2019.
- [13] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 512-519.
- [14] S. Ibrahim, N. A. Zulkifli, N. Sabri, A. A. Shari, and M. R. Mohd, "Rice grain classification using multi-class support vector machine (SVM)," vol. 8, no. 3, pp. 215-220, 2019.
- [15] N. A. Muhammad, A. A. Nasir, Z. Ibrahim, and N. Sabri, "Evaluation of CNN, alexnet and GoogleNet for fruit recognition," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 2, pp. 468-475, 2018.
- [16] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and F. F. Li, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725-1732.
- [17] X. Wang, A. Mohd Ali, and P. Angelov, "Gender and Age Classification of Human Faces for Automatic Detection of Anomalous Human Behaviour," in *International Conference on Cybernetics (CYBCONF 2017)*, 2017, pp. 1-6.
- [18] A. M. Ali and P. Angelov, "Anomalous behaviour detection based on heterogeneous data and data fusion," *Soft Comput.*, no. Dhar 2013, 2018.
- [19] N. A. Ali, A. R. Syafeeza, A. S. Jaafar, M. K. M. F. Alif, and N. A. Ali, "Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm," vol. 9, no. 1, pp. 91-99, 2020.
- [20] L. Lei and K. She, "Identity Vector Extraction by Perceptual Wavelet Packet Entropy and Convolutional Neural Network," *Entropy*, vol. 20, 2018.
- [21] W. Guan, "Performance Optimization of Speech Recognition System with Deep Neural Network Model," *Opt. Mem. Neural Networks*, vol. 27, no. 4, pp. 272-282, 2018.
- [22] H. Rai and A. Yadav, "Iris recognition using combined support vector machine and Hamming distance approach," *Expert Syst. Appl.*, vol. 41, no. 2, pp. 588-593, 2014.
- [23] R. Li, F. Lee, Y. Yan, and Q. Chen, "Face Recognition Using Vector Quantization Histogram and Support Vector Machine Classifier," in *International Conference on Artificial Intelligence: Techniques and Applications (AITA 2016)*, 2016, no. Aita, pp. 144-149.
- [24] F. Luo, C. Li, and Z. Cao, "Affective-feature-based sentiment analysis using SVM classifier," in *2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2016, pp. 276-281.
- [25] M. Ahmad, S. Aftab, and I. Ali, "Sentiment Analysis of Tweets using SVM," in *International Journal of Computer Applications*, 2017, pp. 25-29.
- [26] M. Usha and P. Kavitha, "Anomaly based intrusion detection for 802.11 networks with optimal features using SVM classifier," *Wirel. Networks*, vol. 23, no. 8, pp. 2431-2446, 2017.
- [27] J. Gu, L. Wang, H. Wang, and S. Wang, "A novel approach to intrusion detection using SVM ensemble with feature augmentation," *Comput. Secur.*, vol. 86, pp. 53-62, 2019.
- [28] N. A. S. Selvakumari and V. Radha, "A Voice Activity Detector using SVM and Naïve Bayes Classification Algorithm," in *International Conference on Signal Processing and Communication (ICSPC'17)*, 2017.
- [29] W. Astuti and E. B. W. Riyandwita, "Intelligent automatic starting engine based on voice recognition system," *Proc. - 14th IEEE Student Conf. Res. Dev. Adv. Technol. SCORED 2016*, pp. 1-5, 2017.
- [30] Harvianto, L. Ashianti, Jupiter, and S. Junaedi, "Analysis and Voice Recognition in Indonesia Language using MFCC and SVM Method," *ComTech*, vol. 7, no. 2, pp. 131-139, 2016.
- [31] S. T. S and C. Lingam, "Speaker based Language Independent Isolated Speech Recognition System," in *2015 International Conference on Communication, Information & Computing Technology (ICCICT)*, pp. 1-7, 2015.

BIOGRAPHIES OF AUTHORS

Mohammad Shahrul Izham Bin Sharifuddin received his Diploma in Electronic Engineering (Robotic and Automation) from MARA Japan Industrial Institute and Bachelor of Information Technology (hons) Intelligent Systems Engineering from Universiti Teknologi MARA in 2015 and 2019 respectively. Currently, he is a freelance developer in machine learning and internet of things (IoT) applications.



Dr. Sharifalillah Nordin received her Bachelor of Information Technology in 2001 from Universiti Utara Malaysia (UUM), Master of Science (Internet Computing) in 2003 from University of Surrey, UK, and PhD in Bioinformatics from Universiti Malaya (UM) in 2010. She is currently a Senior Lecturer at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Shah Alam. She has been an academician in UiTM since 2009. She dedicates herself to university teaching and conducting research. Her research fields include biodiversity informatics, knowledge engineering, and artificial intelligence.



Dr Azliza Mohd Ali received both first and master degree from Universiti Utara Malaysia (UUM) in Bachelor of Information Technology (2001) and Master of Science (Intelligent Knowledge-Based System (2003). He joined Universiti Teknologi MARA (UiTM) as a lecturer in 2004 and received a PhD in Computer Science from Lancaster University, UK. She dedicates herself to university teaching and conducting research. Currently her research interest on anomaly detection, data mining, machine learning and knowledge-based systems.